LA-UR-21-27531
Approved for public release; distribution is unlimited.

Title:	A Machine Learning aided hierarchical screening strategy for materials discovery
Author(s):	Talapatra, Anjana Anu
Intended for:	Slides to accompany tutorial (already approved) on nanoHUB.org
Issued:	2021-07-30

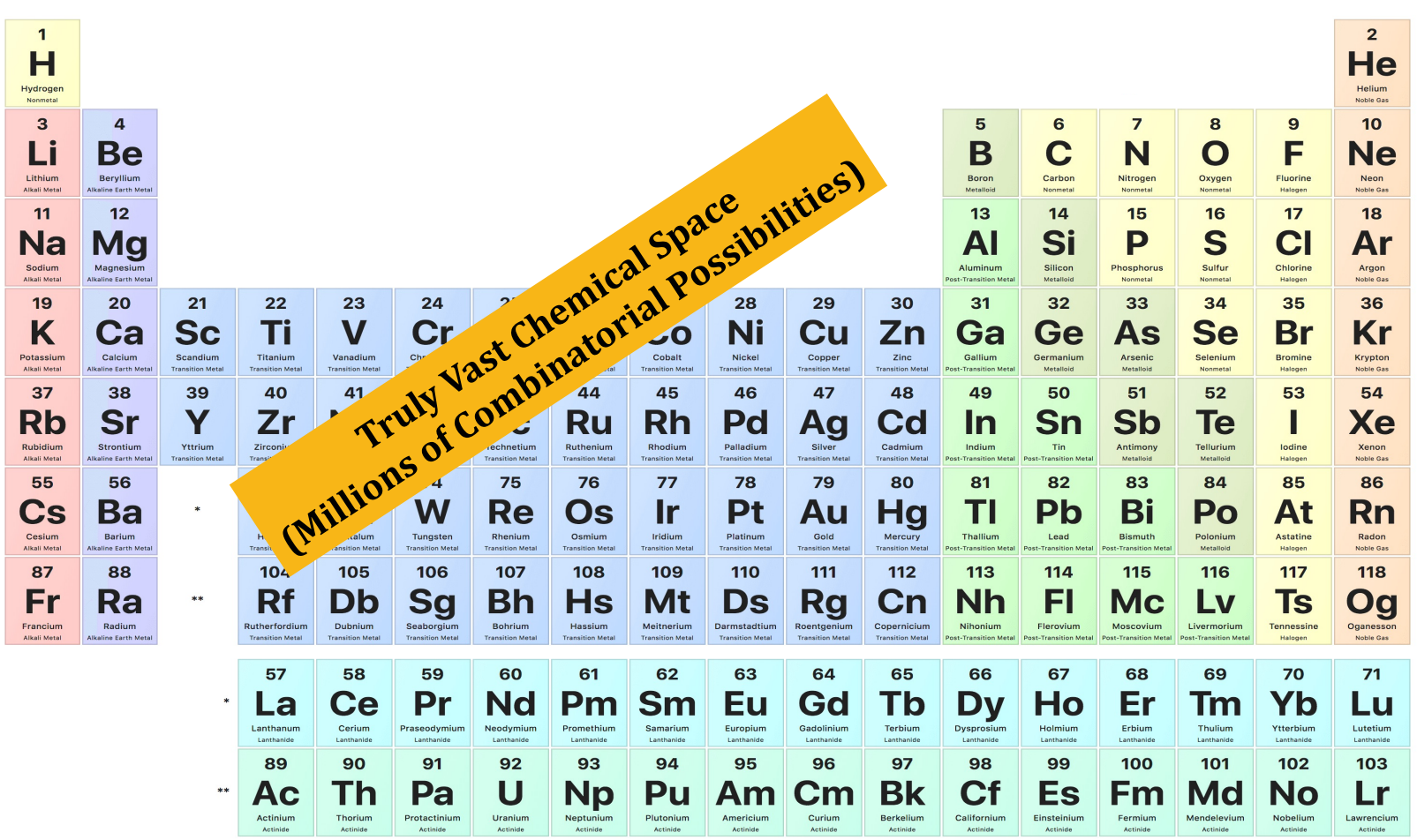
A Machine Learning aided hierarchical screening strategy for materials discovery

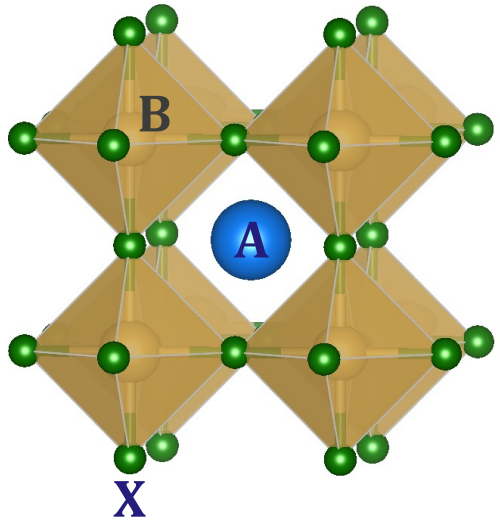
Anjana Talapatra
Director's Postdoctoral Fellow
Materials Science & Technology Division
Los Alamos National Laboratory

21st July, 2021

Managed by Triad National Security, LLC for the U.S. Department of Energy's NNSA

Discovery and Design of Novel wide band gap Oxide Perovskites

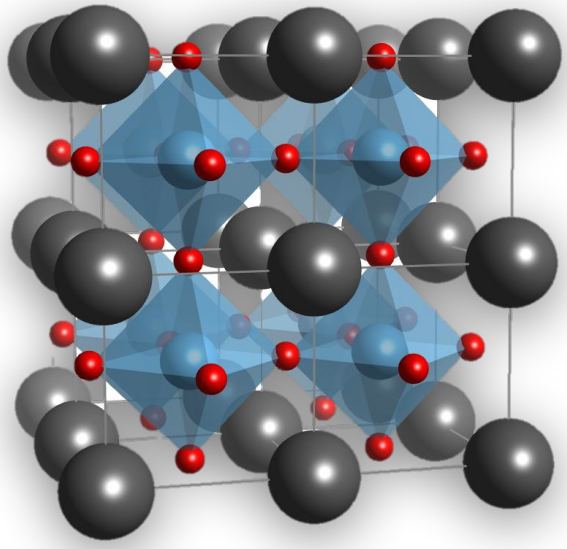




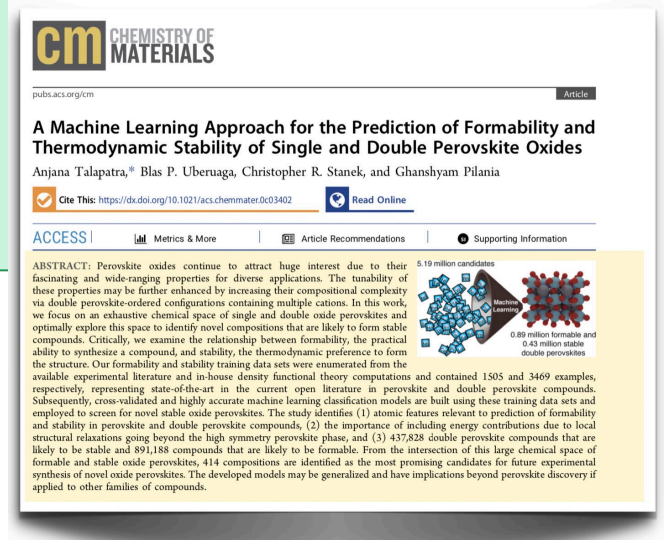
Objective:

Down select from millions of potential compounds to a relatively small and tractable set of promising wide band gap oxide perovskites

Oxide Perovskites



- The perovskite structure can accommodate 90% of the metallic ions in the periodic table
- Amenable to band gap tuning
- Exhibit fascinating electrical and magnetic properties:
  - piezoelectricity, optical properties,
  - high-temperature superconductivity,
  - ferroelectricity, magneto-strictive effects



CHEMISTRY OF MATERIALS

pubs.acs.org/cm Article

A Machine Learning Approach for the Prediction of Formability and Thermodynamic Stability of Single and Double Perovskite Oxides

Anjana Talapatra,* Blas P. Uberuaga, Christopher R. Stanek, and Ghanshyam Pilania

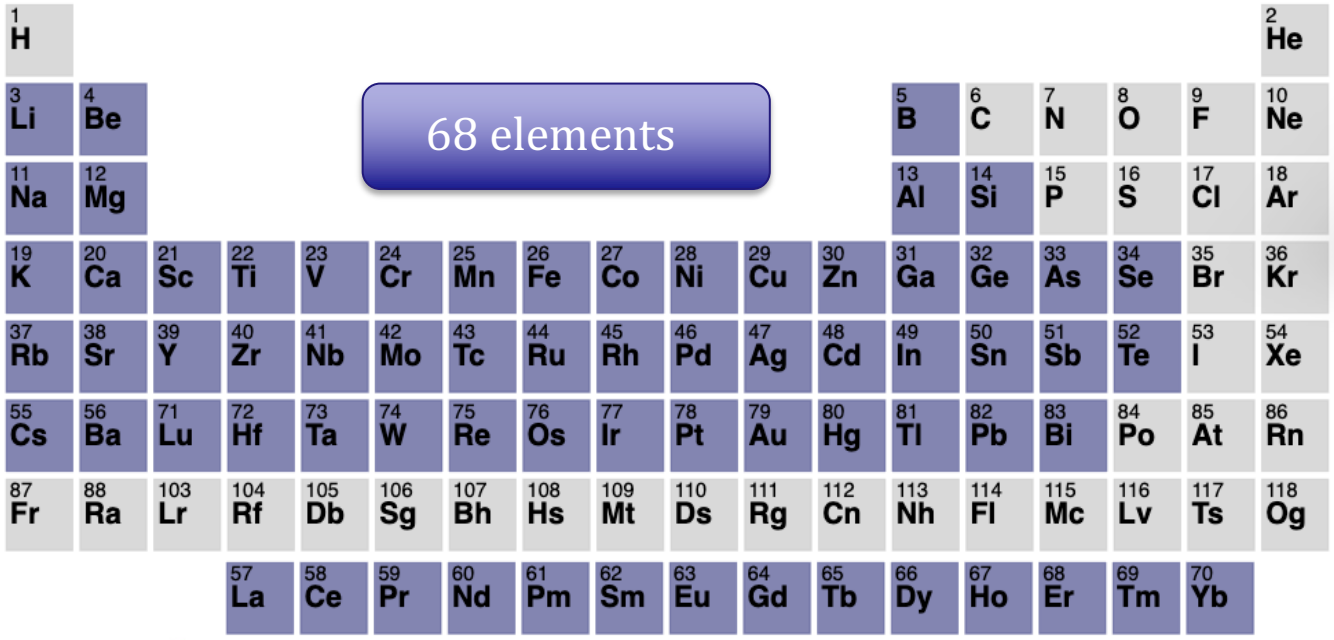
Cite This: https://dx.doi.org/10.1021/acs.chemmater.0c03402 Read Online

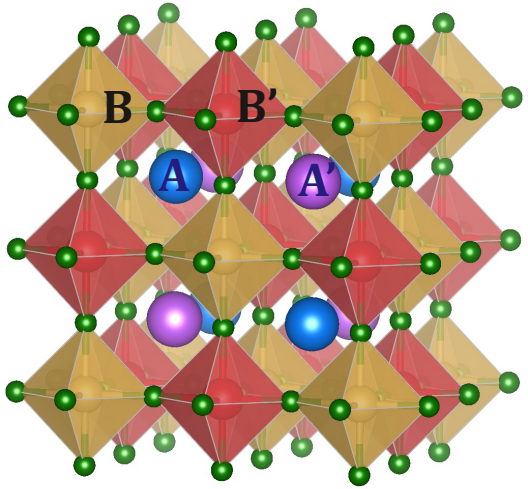
ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Perovskite oxides continue to attract huge interest due to their fascinating and wide-ranging properties for diverse applications. The tunability of these properties may be further enhanced by increasing their compositional complexity via double perovskite-ordered configurations containing multiple cations. In this work, we focus on an exhaustive chemical space of single and double oxide perovskites and optimally explore this space to identify novel compositions that are likely to form stable compounds. Critically, we examine the relationship between formability, the practical ability to synthesize a compound, and stability, the thermodynamic preference to form the structure. Our formability and stability training data sets were enumerated from the available experimental literature and in-house density functional theory computations and contained 1505 and 3469 examples, respectively, representing state-of-the-art in the current open literature in perovskite and double perovskite compounds. Subsequently, cross-validated and highly accurate machine learning classification models are built using these training data sets and employed to screen for novel stable oxide perovskites. The study identifies (1) atomic features relevant to prediction of formability and stability in perovskite and double perovskite compounds, (2) the importance of including energy contributions due to local structural relaxations going beyond the high symmetry perovskite phase, and (3) 437,828 double perovskite compounds that are likely to be stable and 891,188 compounds that are likely to be formable. From the intersection of this large chemical space of formable and stable oxide perovskites, 414 compositions are identified as the most promising candidates for future experimental synthesis of novel oxide perovskites. The developed models may be generalized and have implications beyond perovskite discovery if applied to other families of compounds.

Talapatra, Anjana, et al. "A Machine Learning Approach for the Prediction of Formability and Thermodynamic Stability of Single and Double Perovskite Oxides." Chemistry of Materials (2021).

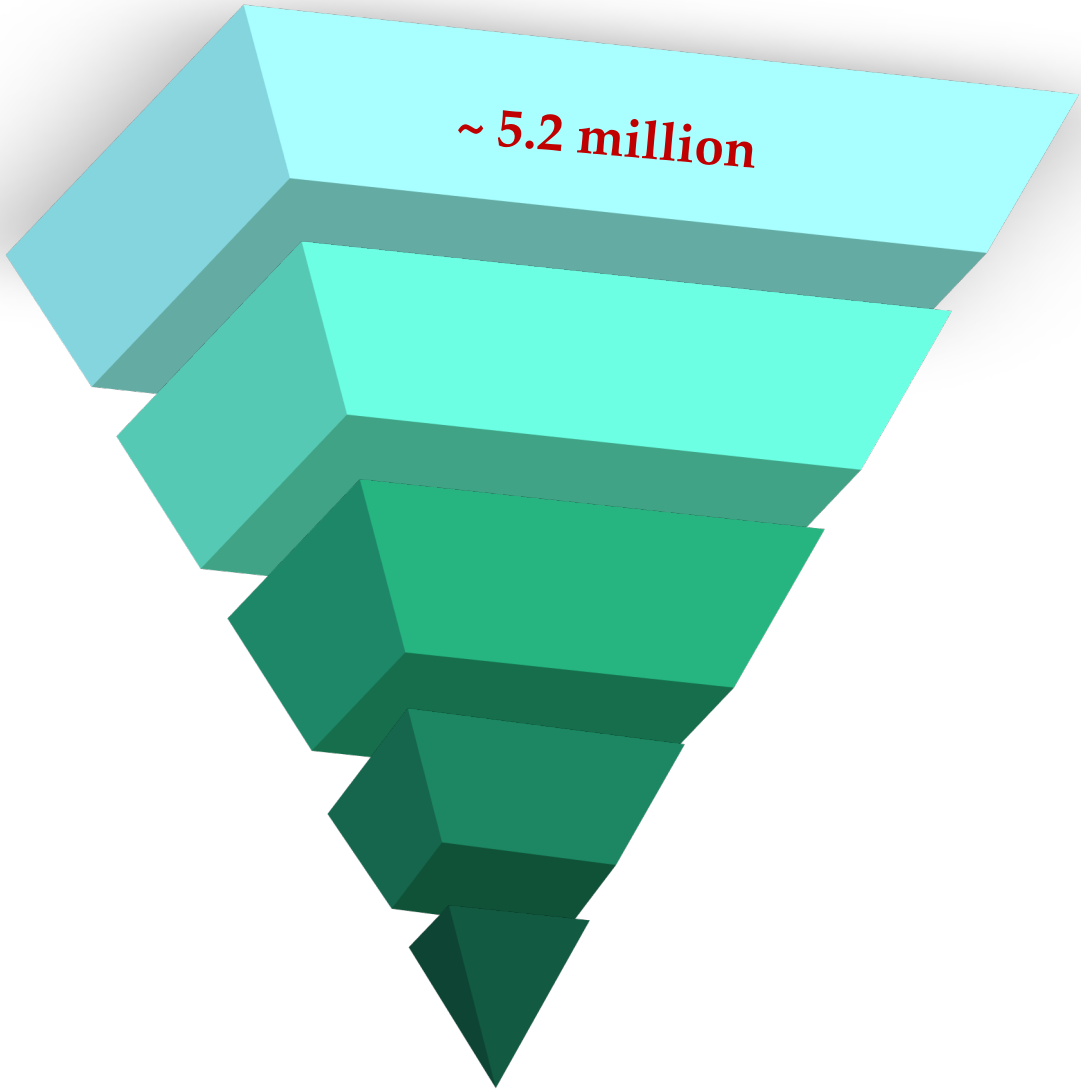
A Strategy for Scintillator Discovery and Optimization





A	A'	B	B'
68	66	67	65

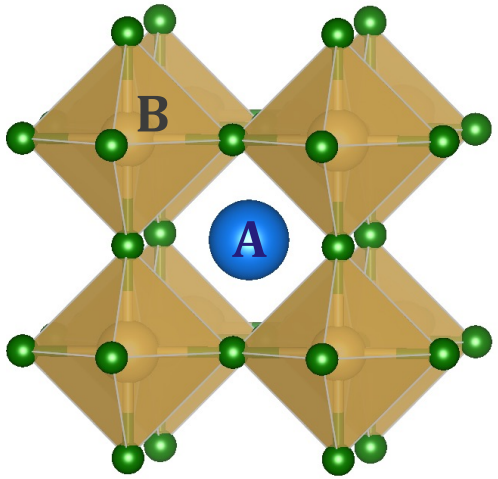
~ 5.2 million
unique combinations



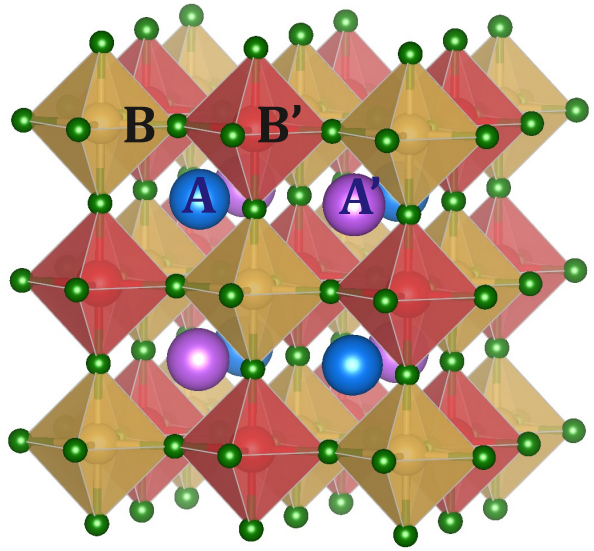
Objective:

Down select from millions of potential compounds to a relatively small and tractable set of promising scintillators

Single perovskite (ABO_3)



Double perovskites



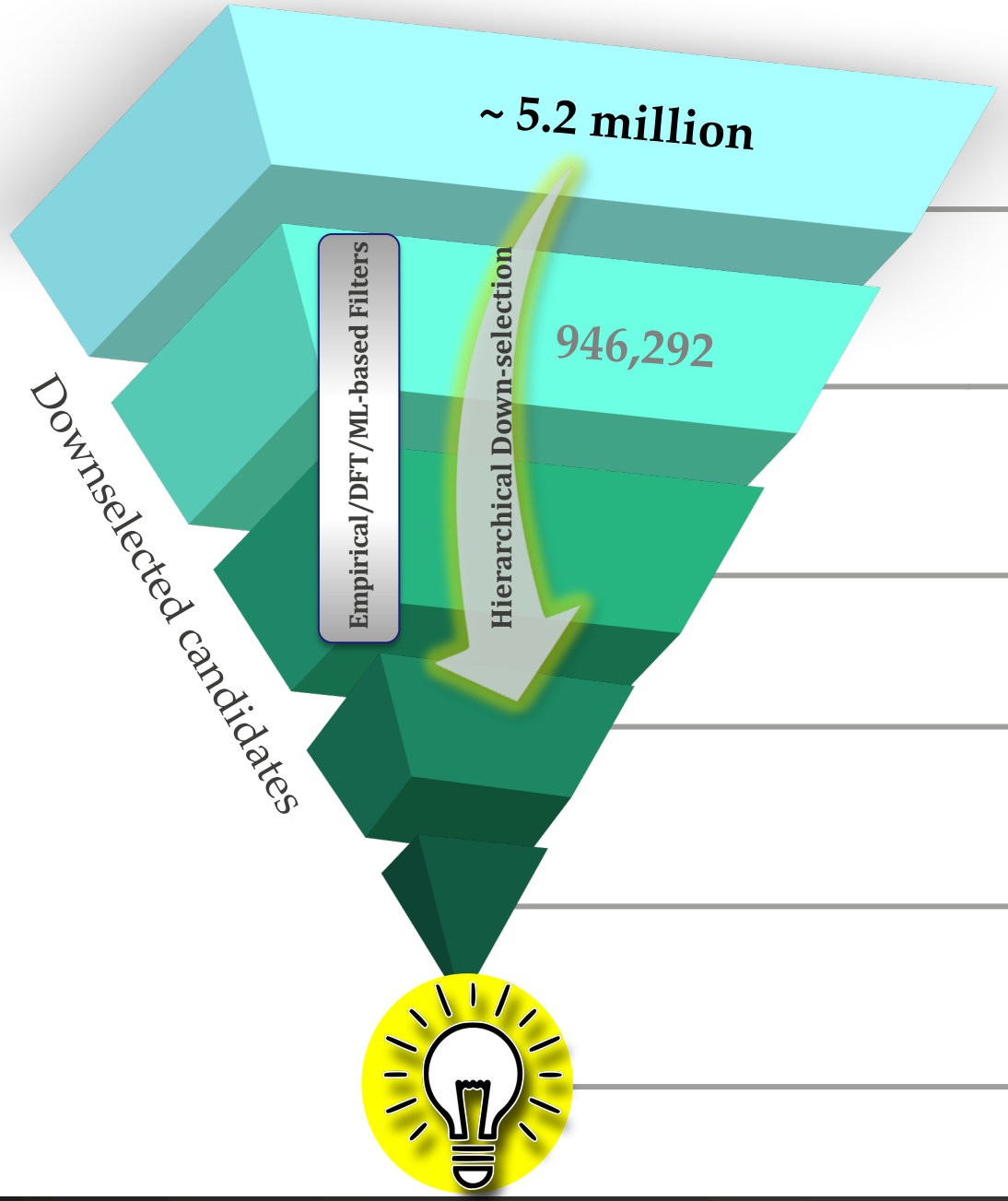
- $AA'B_2O_6$
- $A_2BB'O_6$
- $AA'BB'O_6$

- Assumptions:
  - 50-50 compositions
  - 2 elements per cation site
  - Cubic structures
  - Rocksalt ordering

Why Machine Learning?

- To screen millions of compounds
- Unearth relationships between electronic structure, chemistry, thermodynamic stability, formability and band gap
- We know how to:
  - Calculate thermodynamic stability
  - Calculate approximate band gap
- Very complicated to:
  - Estimate synthesizability

Machine Learning for Efficient Screening



Perovskite domain

Chemically compatible

- Charge neutrality
- Include all feasible combinations of oxidation states
- Cation size (A > B)

Stable + Formable

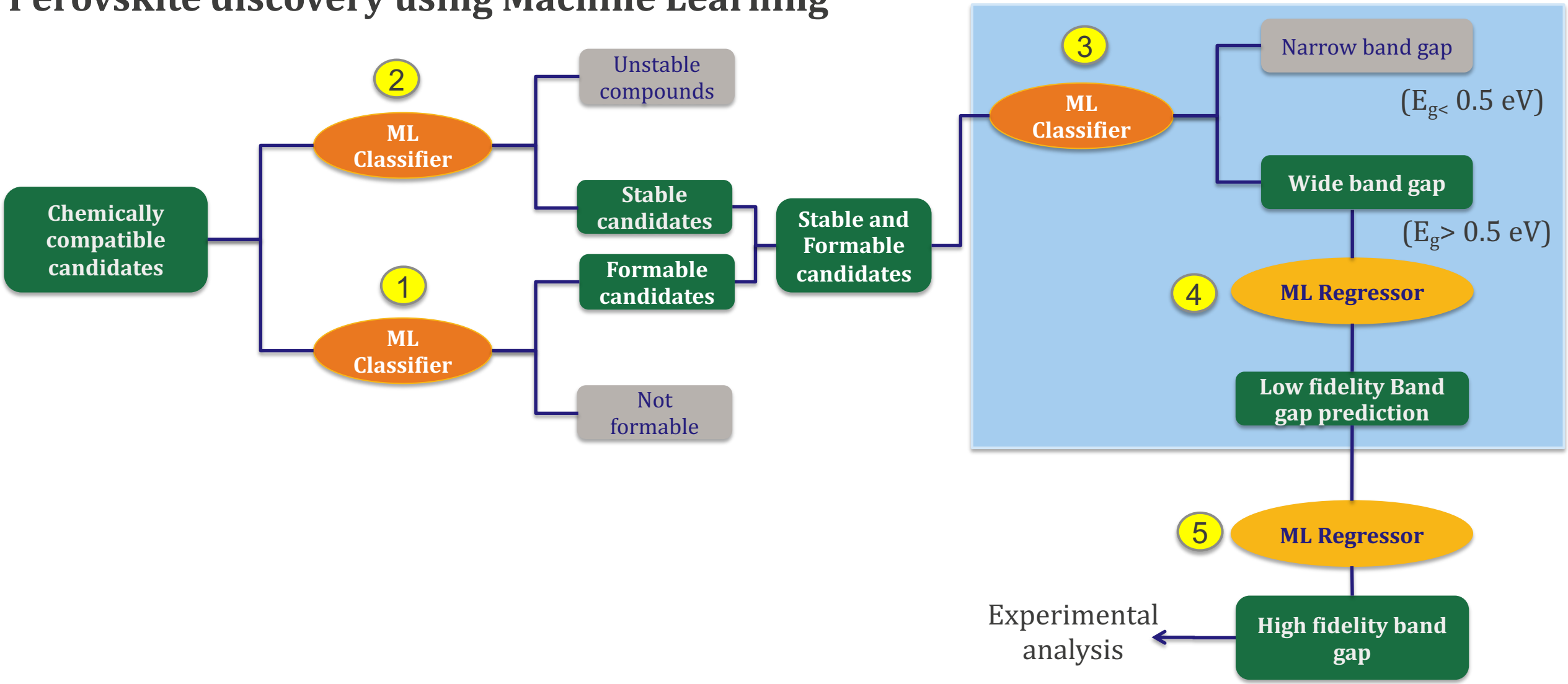
Wide band gap

Down -selected candidates

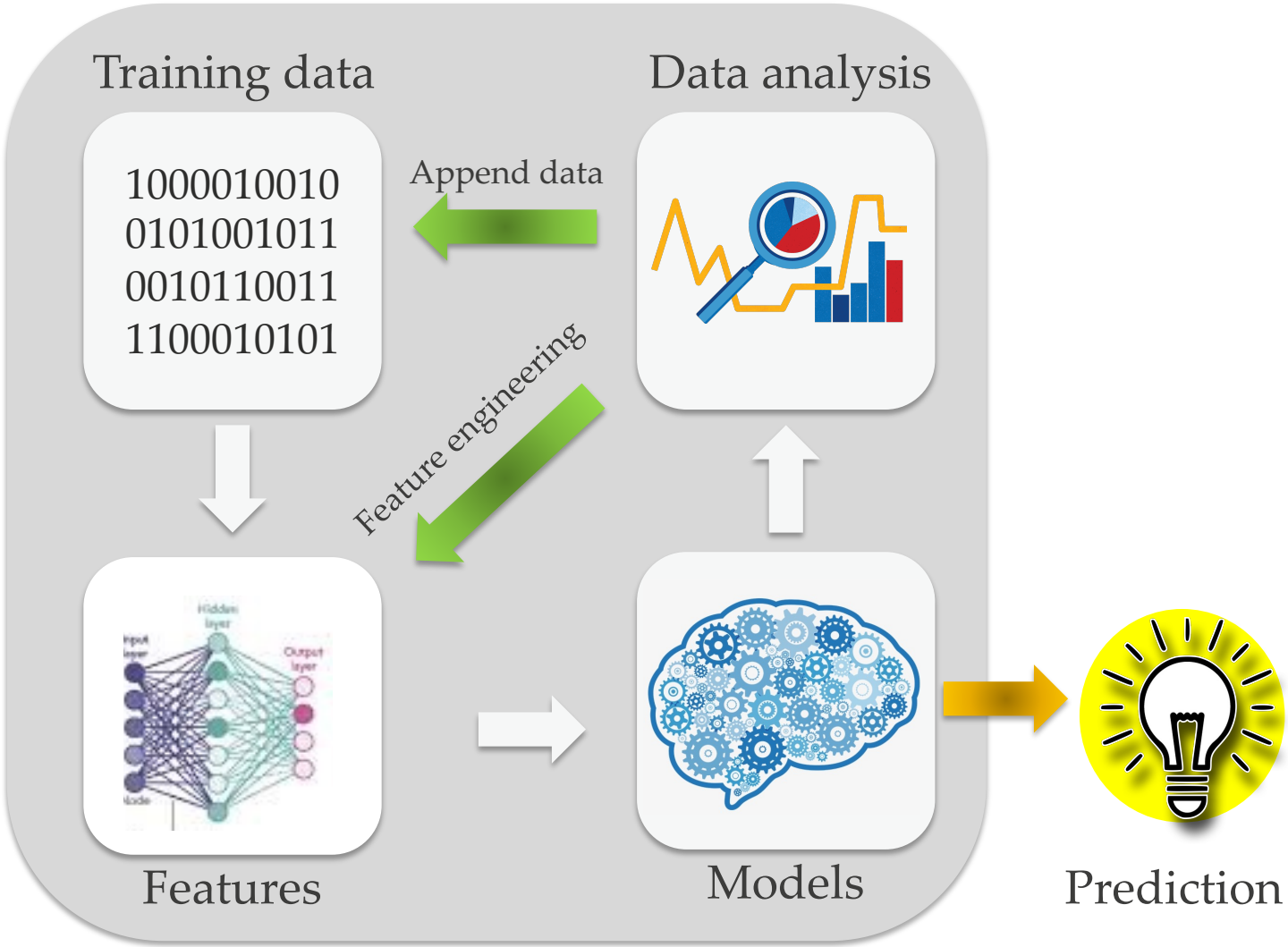
Recommendations for application-specific in-depth analysis

Machine learning Models

Perovskite discovery using Machine Learning

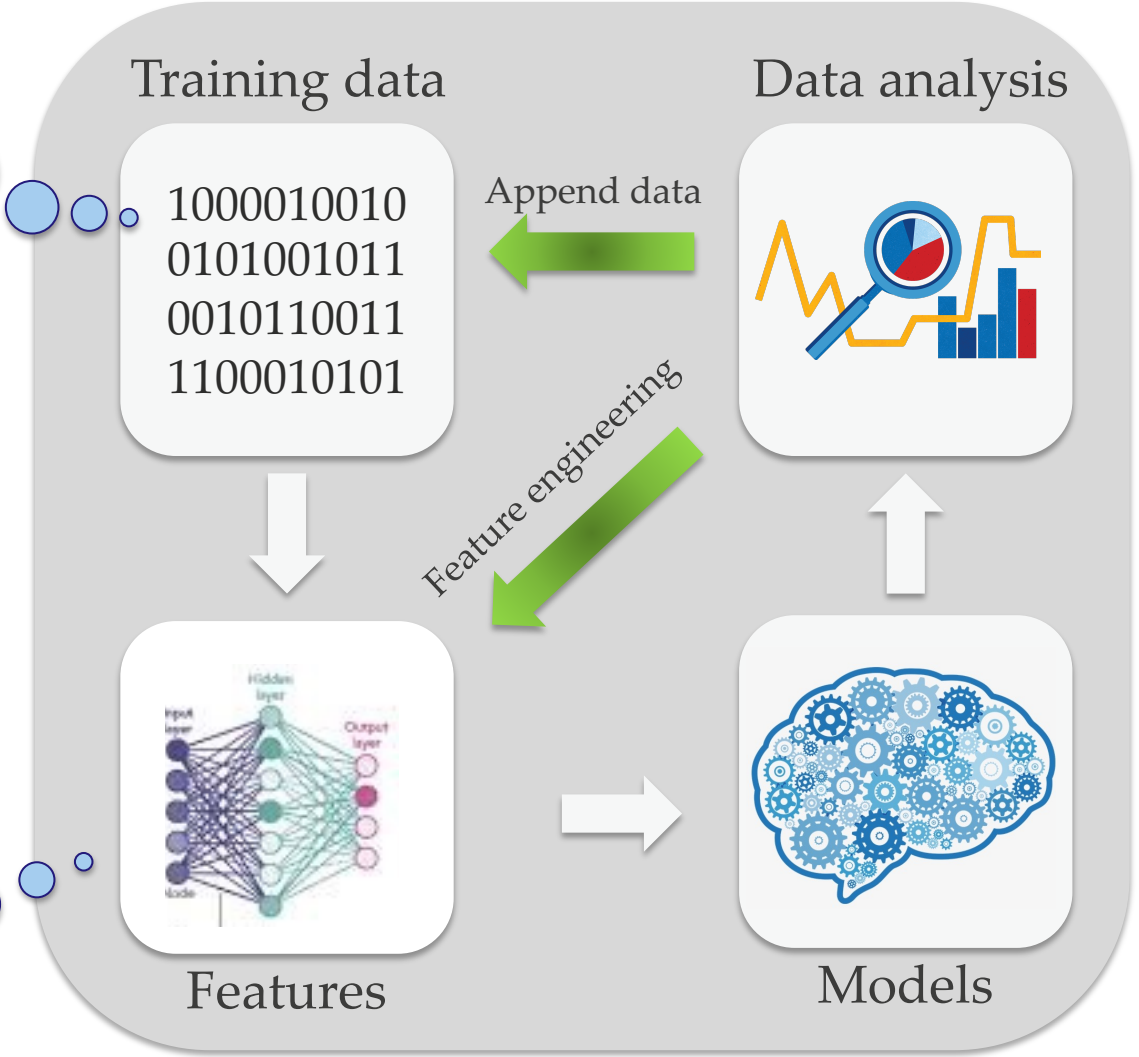


Components of ML infrastructure



Components of ML infrastructure

- Experimental data from literature
- DFT calculations

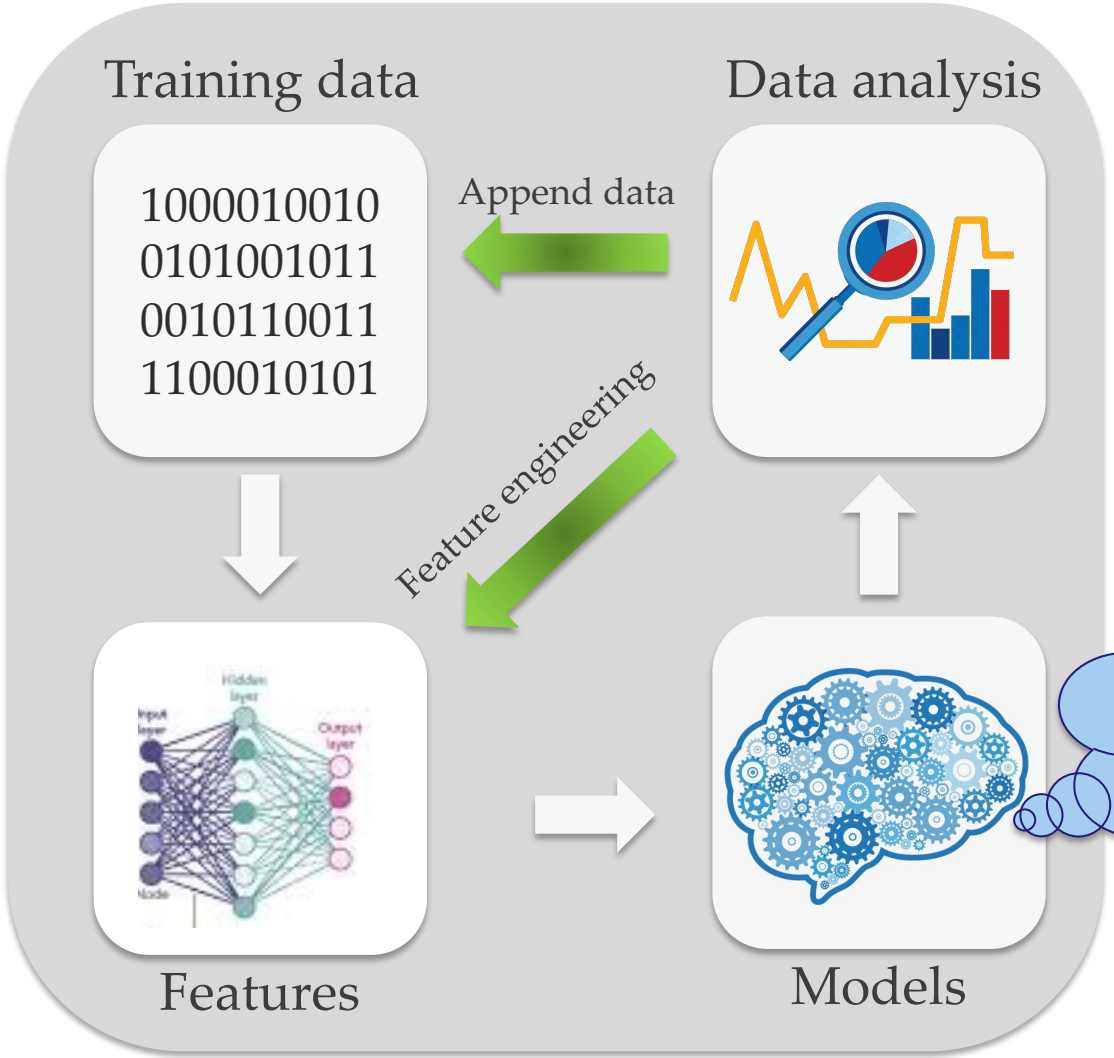


- Elemental features
- Domain knowledge

- Model is as good as your training data
- To increase applicability, ideally use features that are easy to populate
- Double check, triple check source of features and values of features.
- 90 % of bugs can be attributed to mistakes in populating training data
- Reproducibility and consistency

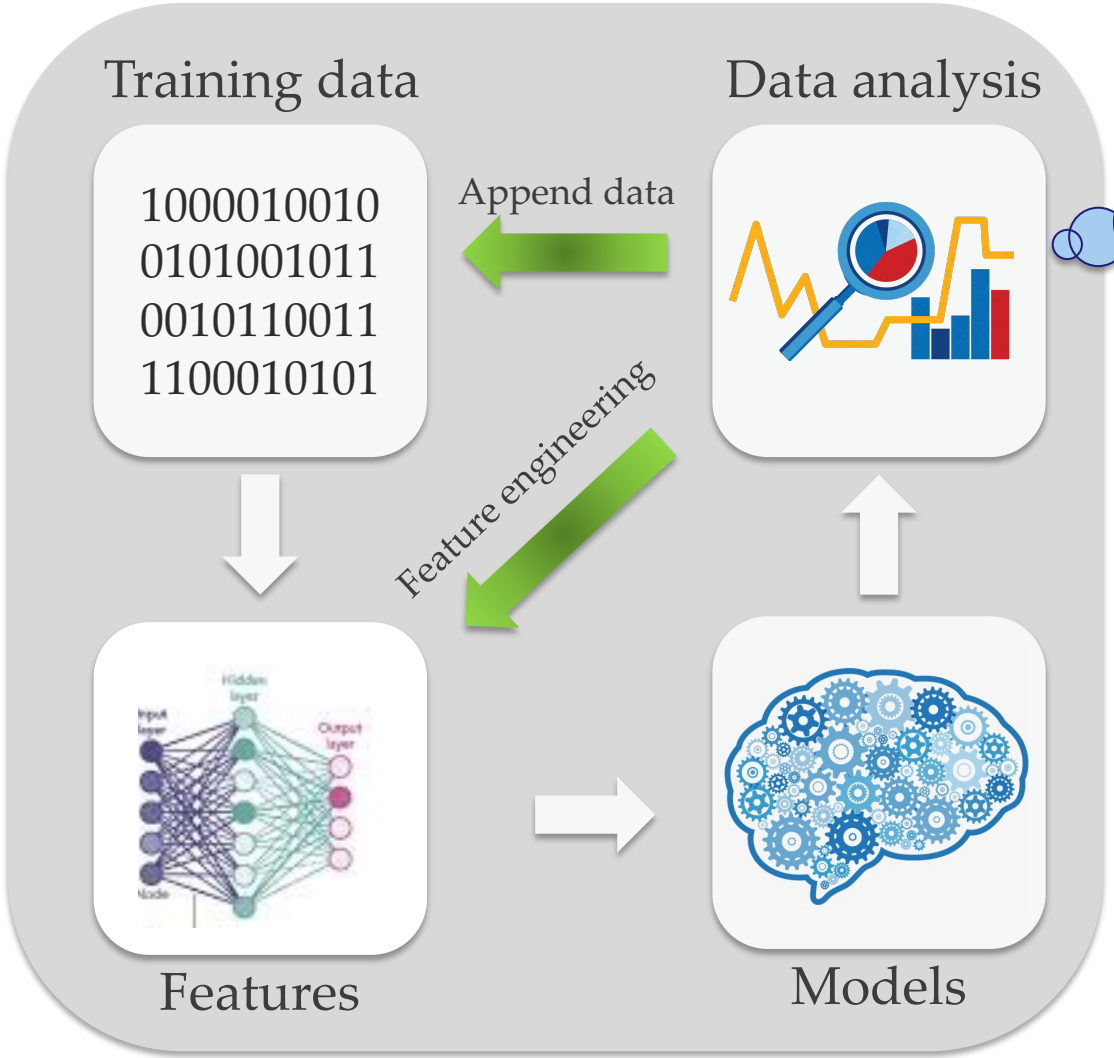
Components of ML infrastructure

- Choose modeling technique wisely
- Baseline comparisons across different models
- Ensure that what you are producing is better than what is available, else use what is available



Components of ML infrastructure

- If possible, increase training data adaptively keeping an eye on performance metrics
- Analyze data and results continuously to ensure it makes sense intuitively
- Avoid data leakage
- If something seems off, it probably is off.



Cross-validation
Feature selection
Testing
Performance curves
Partial dependence plots

Training data

1. Formability classification → Training data: Experimental
   - Training data compiled from literature and experimental databases (ICSD etc)
   - 1505 single and double oxide compositions
     - 1187 perovskites
     - 318 non-perovskites

2. Thermodynamic stability classification → Training data: DFT
   - Criterion:
     - Energy above hull < 50 meV/atom
   - 3271: stable
   - 1881: unstable

3. Wide/narrow band gap classification → Training data: DFT
   - Criterion: Band gap (E_g > 0.5 eV)
     - 1575 : wide band gap
     - 3577: narrow band gap

4. Band gap regression → Training data: DFT
   - 1575 : wide band gap materials

Training data:
- Calculated using DFT
- 5152 compounds
- GGA - PBE formalism

Features: Machine Learning models

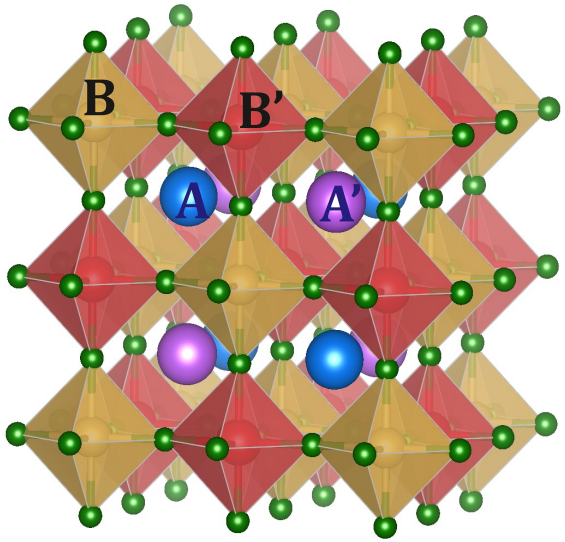
Combination of chemical and structural features		
Pseudopotential radius*	Highest occupied atomic orbital (HOMO)*	Tolerance factor (τ)
Electronegativity*	Lowest occupied atomic Orbital (LUMO)*	Octahedral factor ($\bar{\mu}$)
Electron affinity*	Ionization energy	Mismatch factor ($\Delta\mu$)

* *Element specific*

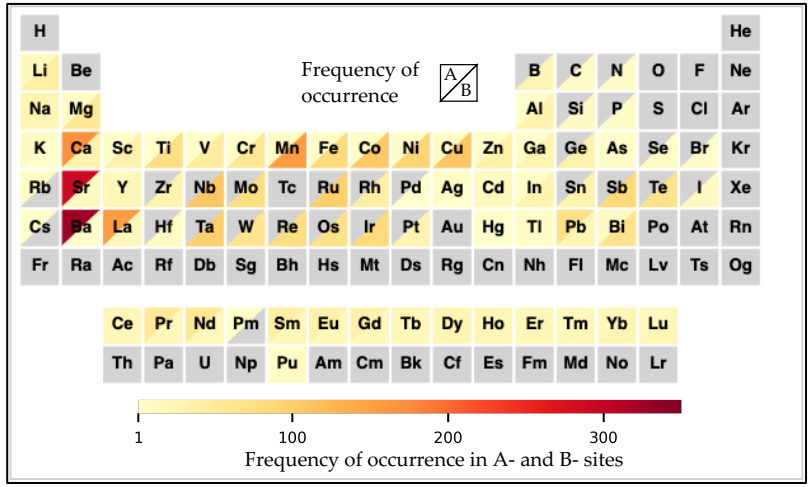
$$\tau = \frac{r_A + r_O}{\sqrt{2}\,(r_B + r_O)}$$

$$\bar{\mu} = \frac{r_B}{r_O}$$

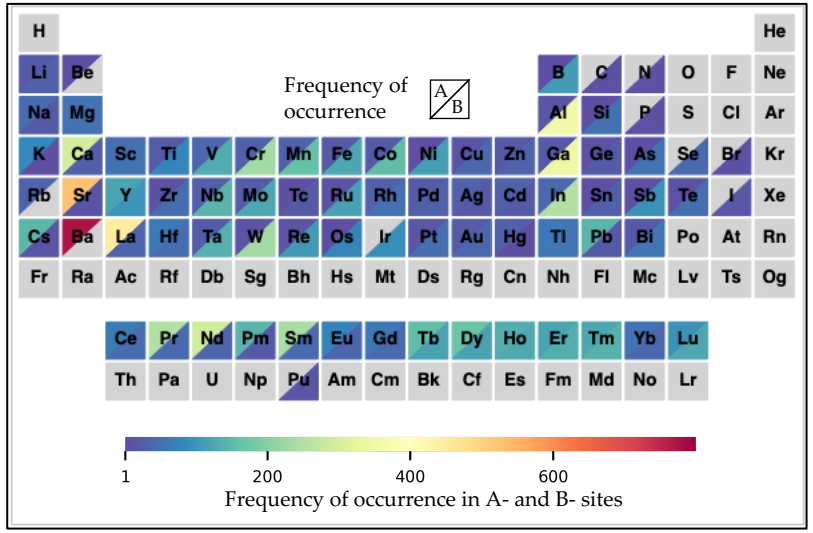
$$\Delta\mu = \frac{|r_{A/B} - r_{A'/B'}|}{r_O}$$



Formability Training dataset



DFT Training dataset



Comparison of perovskite formability and stability

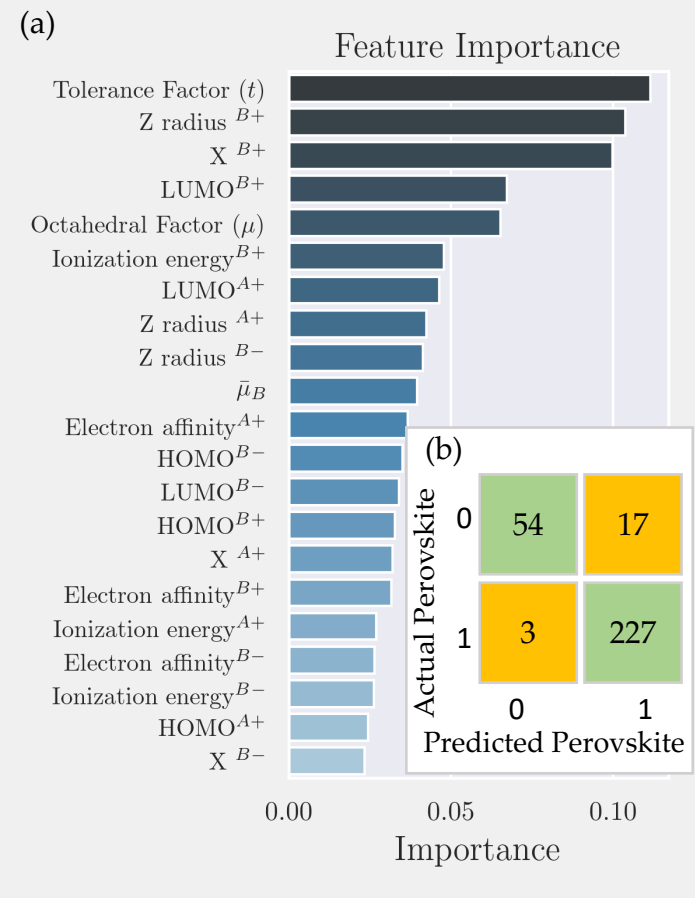
- Formability: ***Ability to experimentally synthesize a model***
- Relies on geometric criteria derived using either ionic radii or bond distances and is a qualitative approach to identifying chemistries that will form perovskites.

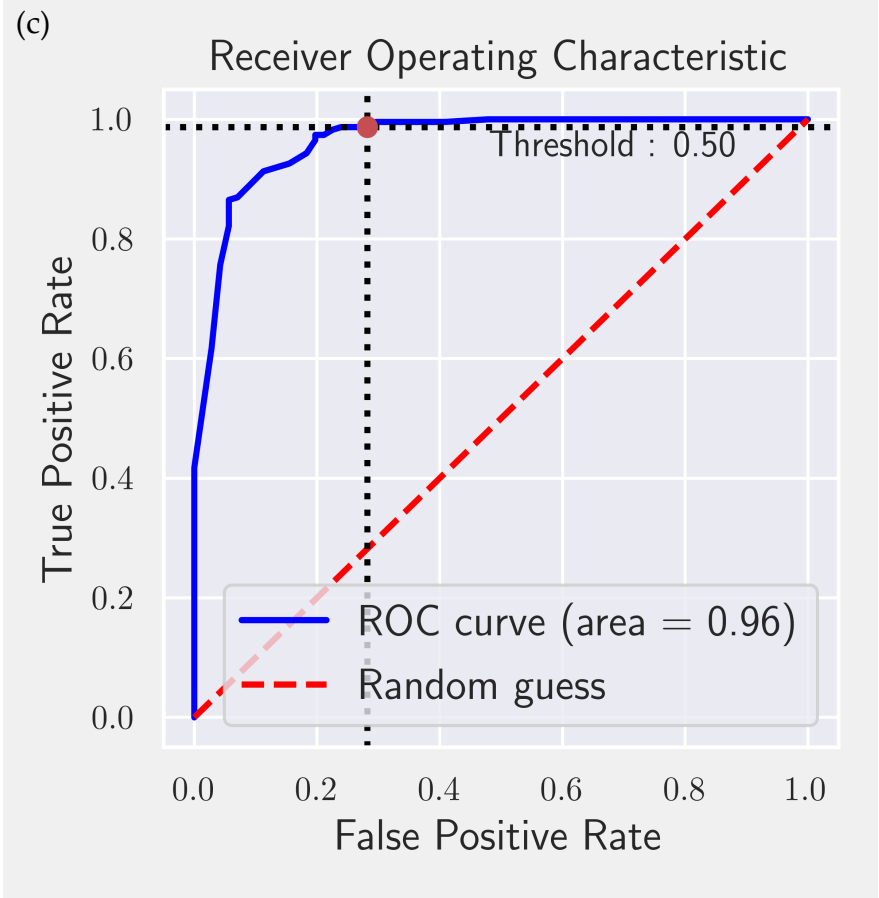
- Stability: ***Thermodynamic preference to form the structure***
- Energy hull construction to determine if the structure is on the convex hull and will stabilize and not decompose.

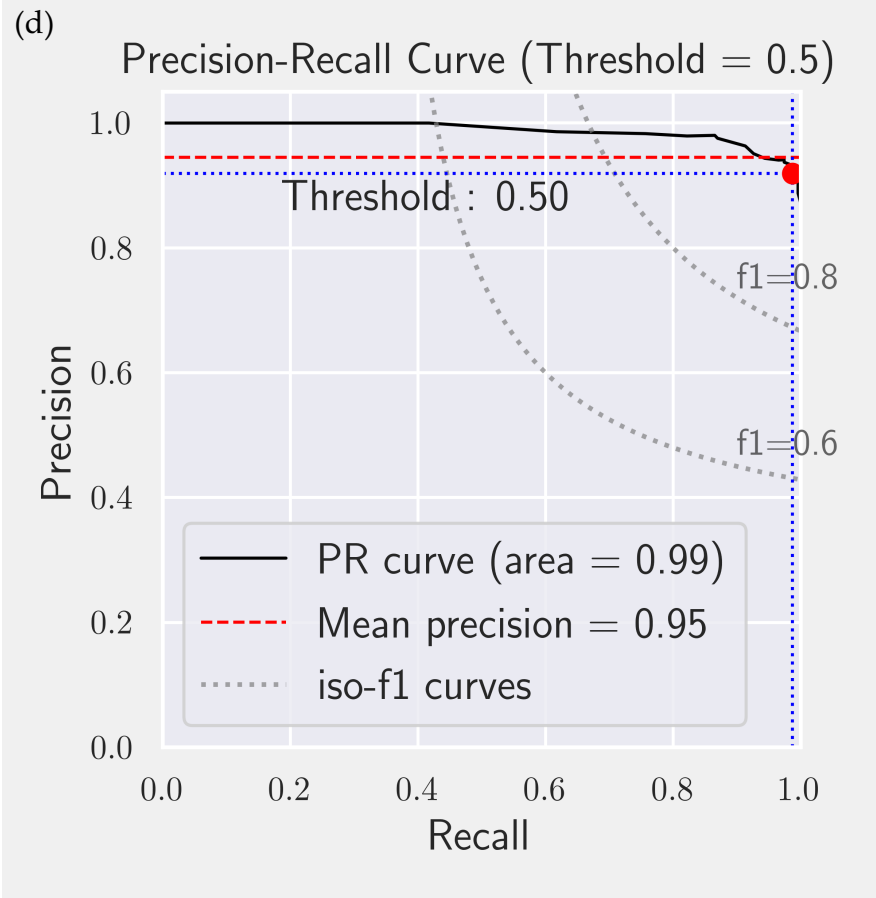
It is not known with certainty whether a formable perovskite is necessarily thermodynamically stable and vice versa.

Are both formability and thermodynamic stability necessary to guarantee the viability of a composition as a perovskite candidate or is one a more robust metric compared to the other?

Formability classification model

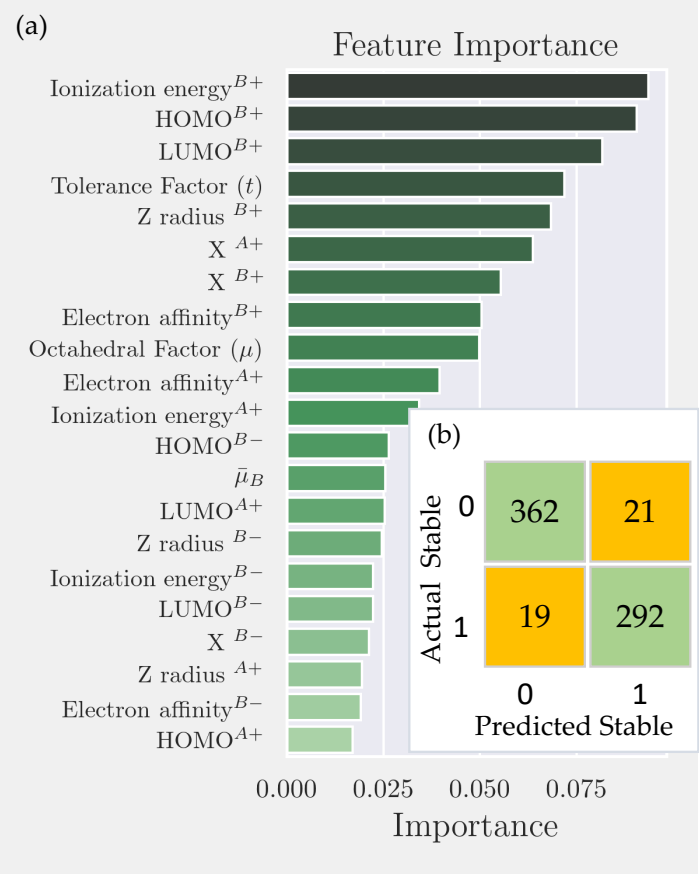


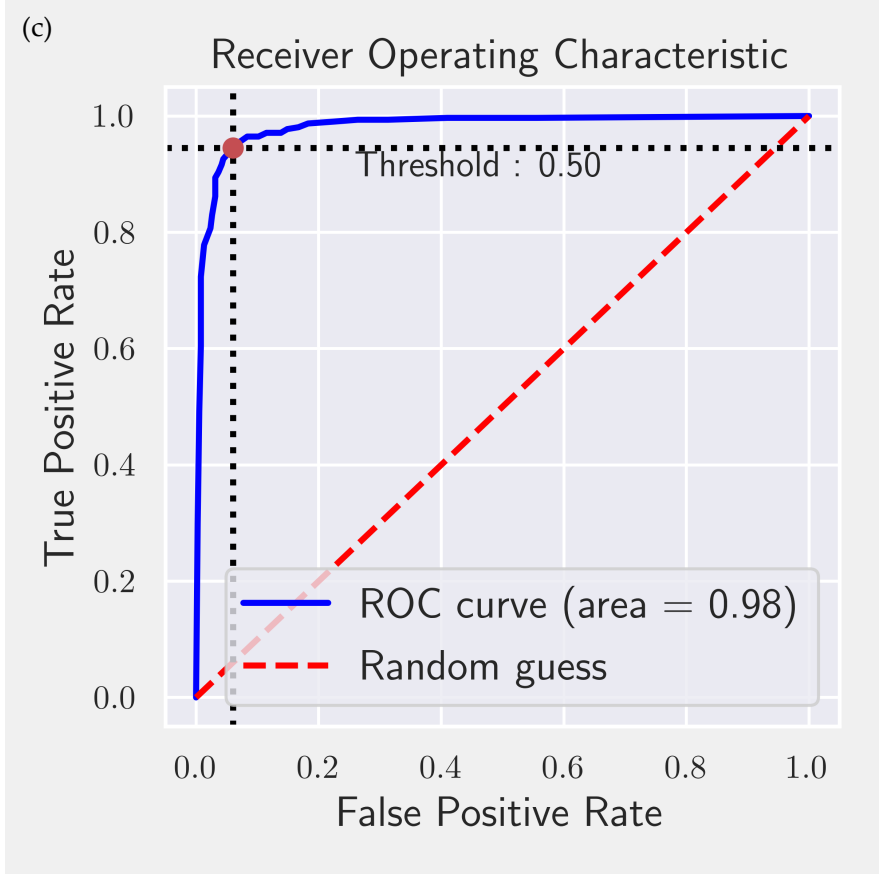


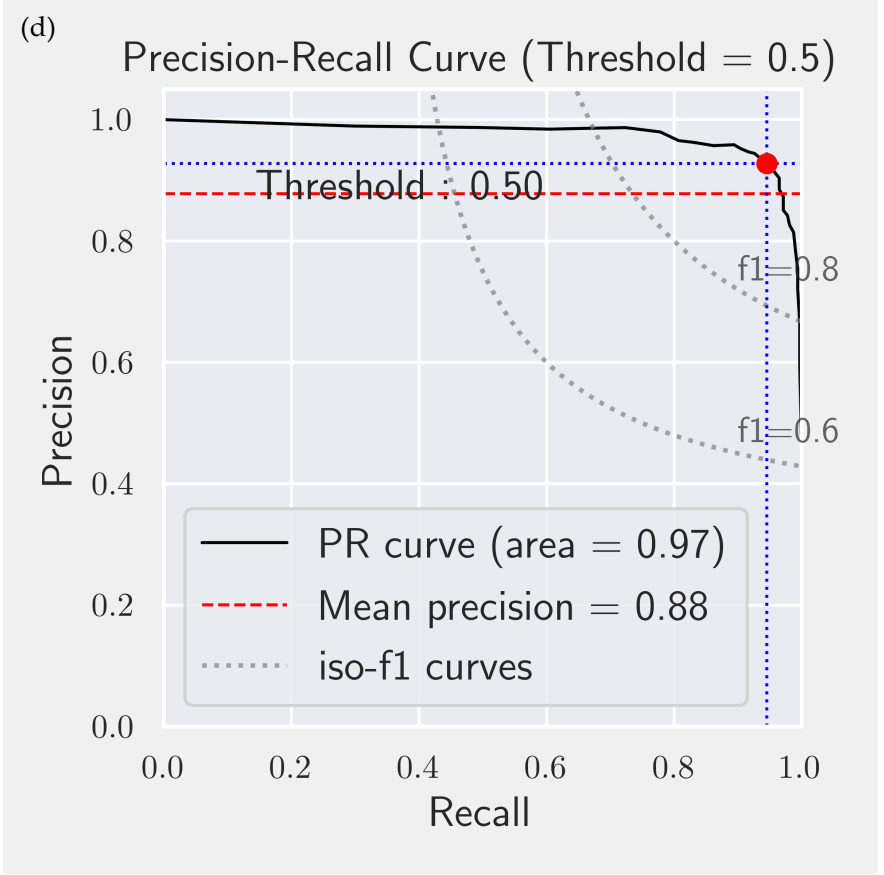


Random forest classification results for perovskite formability. a) Feature importance plot for all the features with non-zero values, b) Confusion matrix, c) Receiver operating characteristic (ROC) curves, and d) Precision-recall curves of the cross-validated random forest classification on test data.

Stability classification model



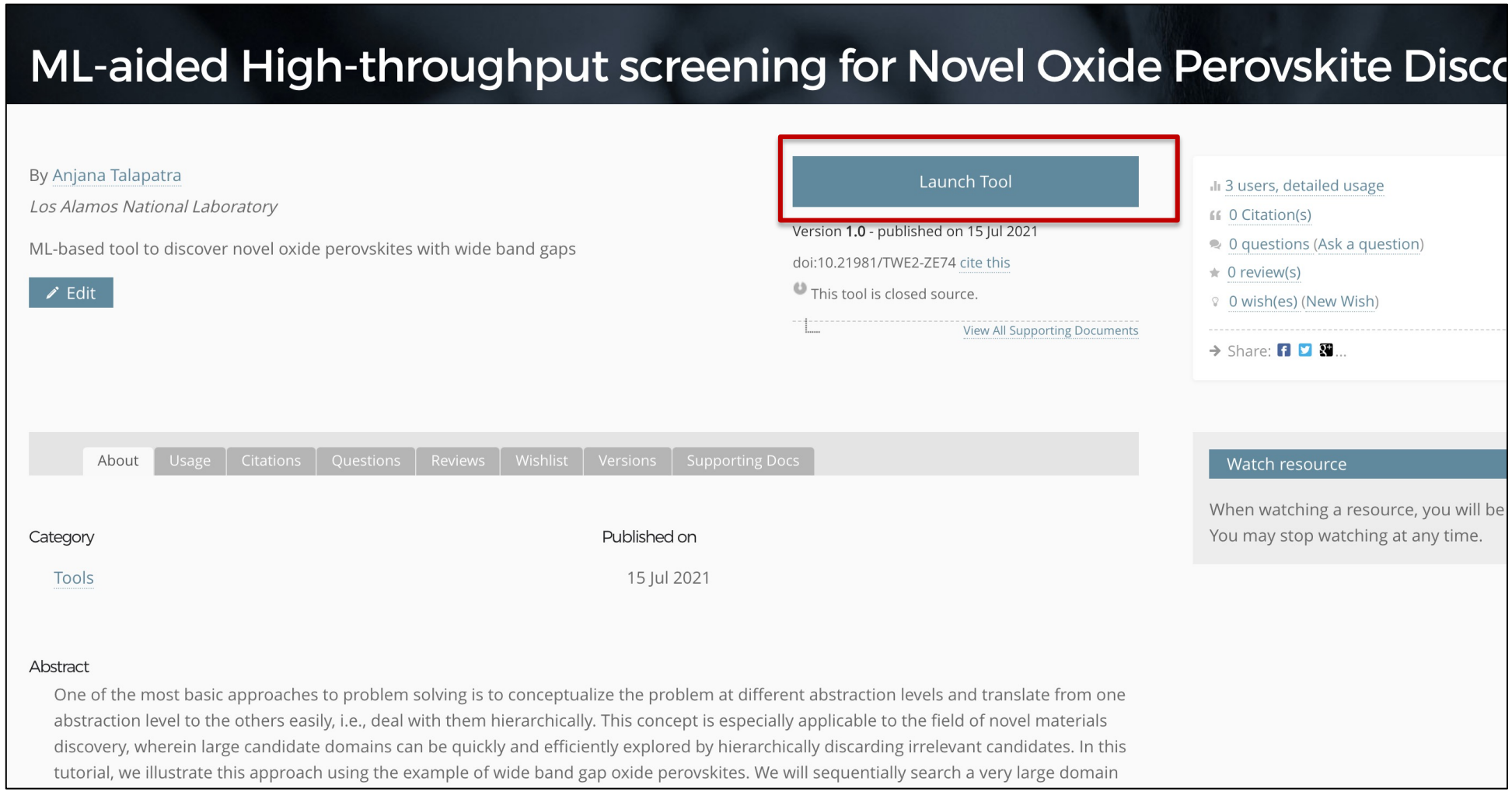




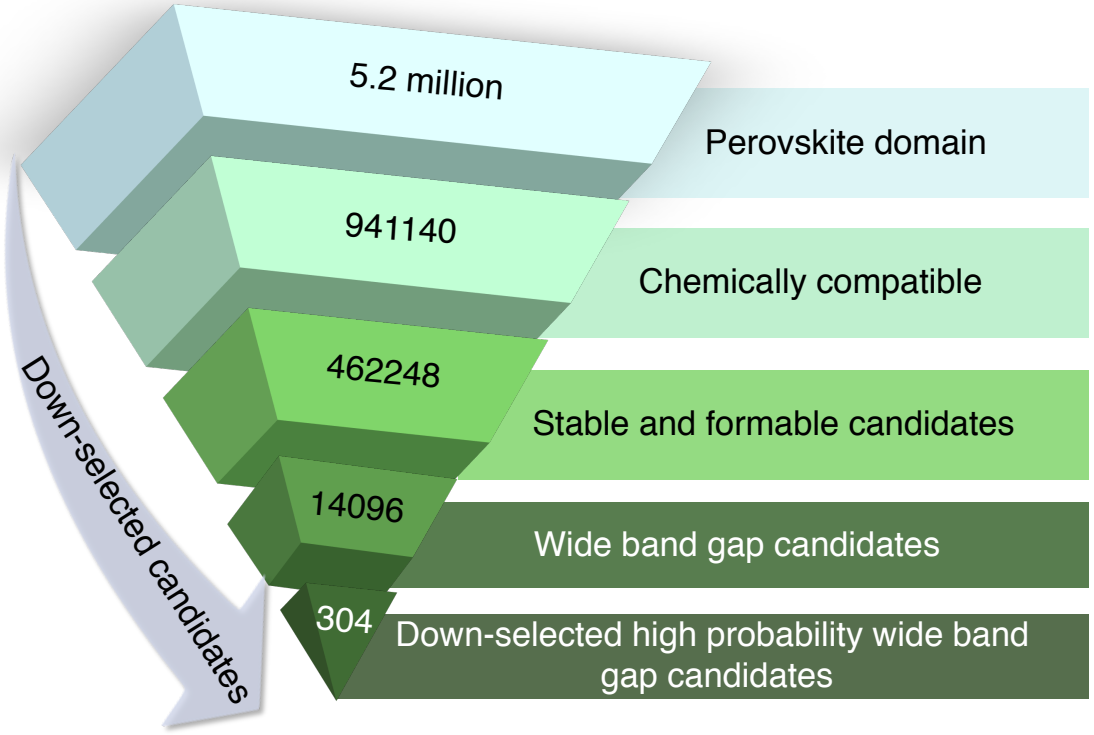
Random forest classification results for perovskite stability. a) Feature importance plot for all the features with non-zero values, b) Confusion matrix, c) Receiver operating characteristic (ROC) curves, and d) Precision-recall curves of the cross-validated random forest classification on test data.

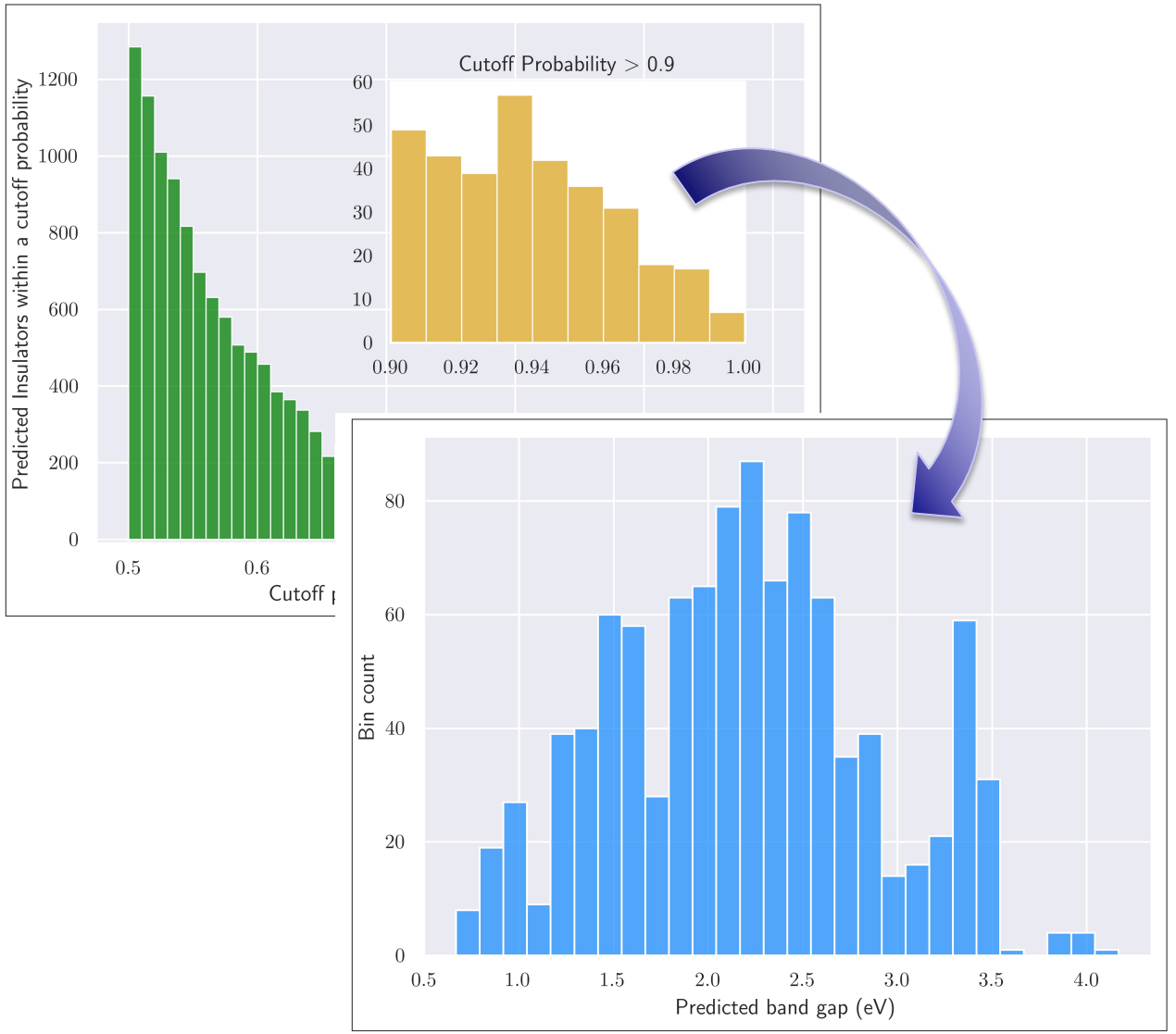
Wide/narrow band gap classification model

https://nanohub.org/tools/perovMLdis

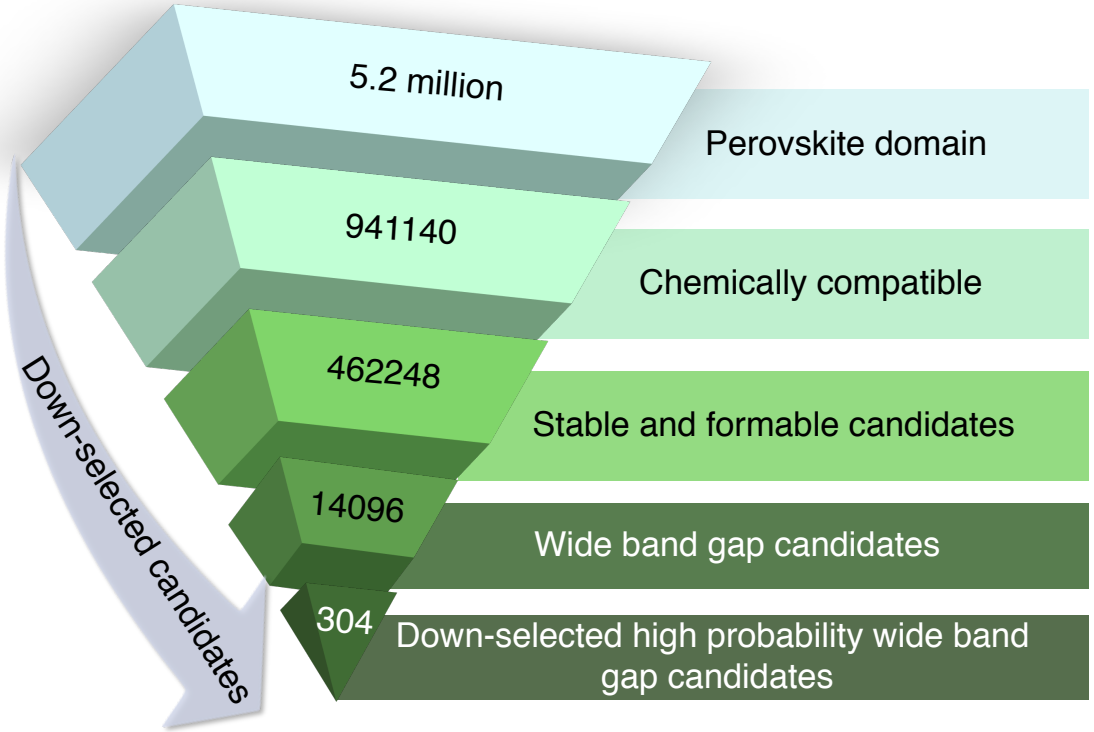


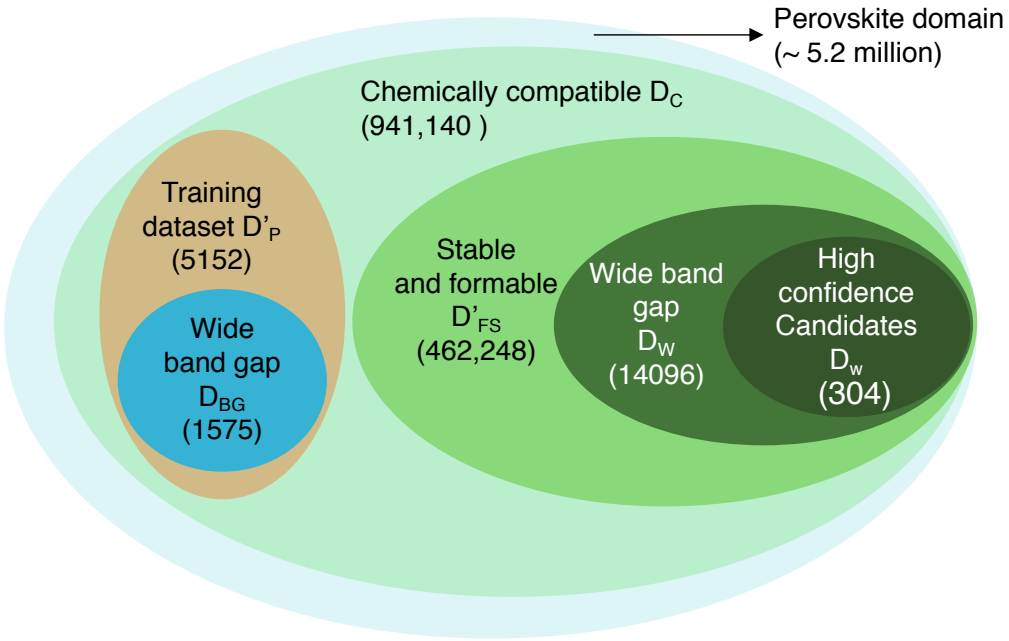
Novel wide bandgap oxide perovskite predictions



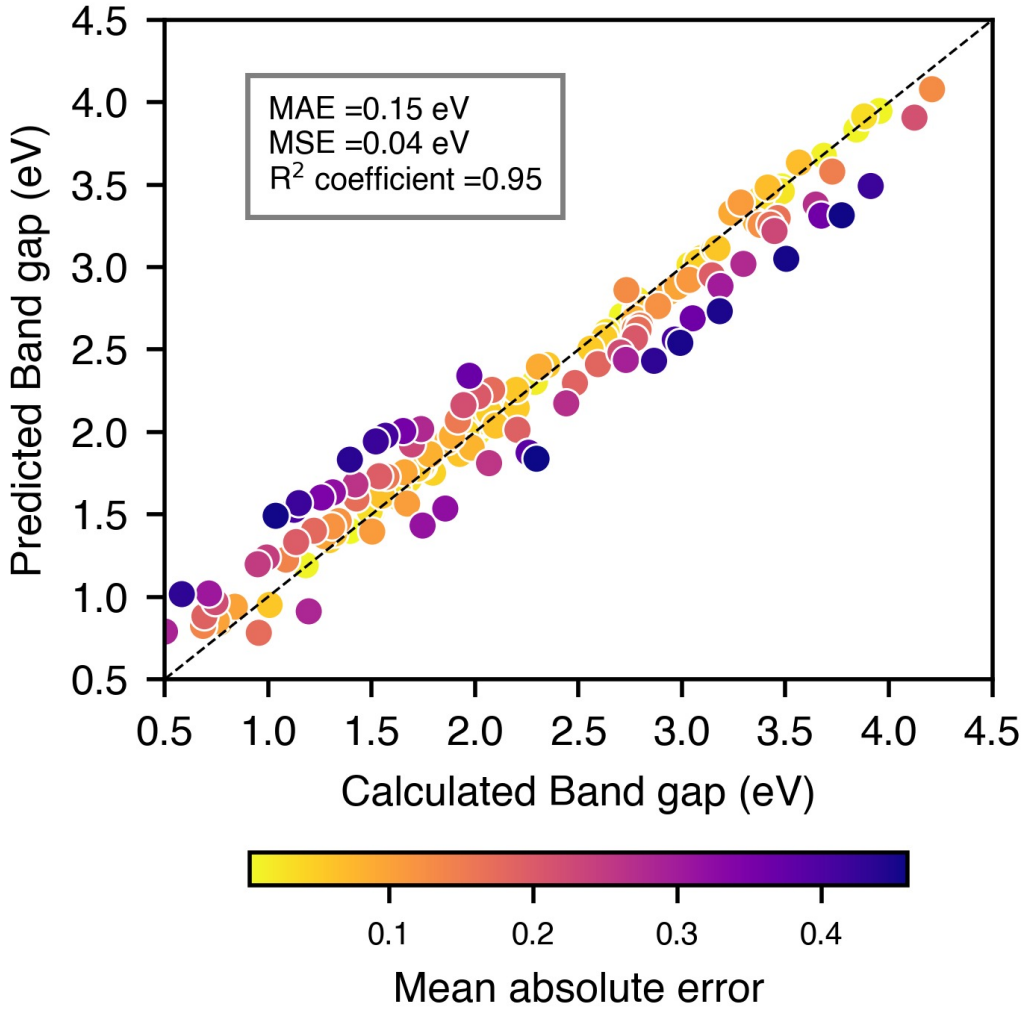


Novel wide bandgap oxide perovskite predictions





Computational confirmation of results



- 150 of the predicted 304 candidates were randomly selected and DFT calculations carried out.
  - Wide band gap ✓
  - Calculated bandgaps accurate with a average MAE = 0.15 eV

Some more suggestions

- Use machine learning only if necessary
- Ensure code is reproducible, with no ad-hoc measures, and all data sources annotated if applicable
- If permissible, have data and codes and scripts publicly available in a repository
- Answer emails from fellow researchers regarding published work and repositories

Thank you!